Energy-induced Explicit quantification for Multi-modality MRI fusion

Xiaoming Qi[1,2], Yuan Zhang[1], Tong Wang[1], Guanyu Yang[1] ✉, Yueming Jin[2] ✉, and Shuo Li[3]

[1] Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications (Southeast University), Ministry of Education, China
yang.list@seu.edu.cn
[2] Department of Biomedical Engineering and Department of Electrical and Computer Engineering, National University of Singapore, Singapore
ymjin@nus.edu.sg
[3] Departments Biomedical Engineering, and Computer and Data Science, Case Western Reserve University

Abstract. Multi-modality magnetic resonance imaging (MRI) is crucial for accurate disease diagnosis and surgical planning by comprehensively analyzing multi-modality information fusion. This fusion is characterized by unique patterns of information aggregation for each disease across modalities, influenced by distinct inter-dependencies and shifts in information flow. Existing fusion methods implicitly identify distinct aggregation patterns for various tasks, indicating the potential for developing a unified and explicit aggregation pattern. In this study, we propose a novel aggregation pattern, Energy-induced Explicit Propagation and Alignment (E^2PA), to explicitly quantify and optimize the properties of multi-modality MRI fusion to adapt to different scenarios. In E^2PA, (1) An energy-guided hierarchical fusion (EHF) uncovers the quantification and optimization of inter-dependencies propagation among multi-modalities by hierarchical same energy among patients. (2) An energy-regularized space alignment (ESA) measures the consistency of information flow in multi-modality aggregation by the alignment on space factorization and energy minimization. Through the extensive experiments on three public multi-modality MRI datasets (with different modality combinations and tasks), the superiority of E^2PA can be demonstrated from the comparison with state-of-the-art methods. Our code is available at https://github.com/JerryQseu/EEPA.

Keywords: Energy model · Multi-modality MRI · Explicit quantification

1 Introduction

Multi-modality magnetic resonance imaging (MRI) fusion offers critical anatomical and functional information for promoting the accuracy and success in disease diagnosis and surgical planning. MRI, being the gold-standard technique for noninvasive tissue characterization, encompasses a range of modalities such as LGE (late gadolinium enhanced), Flair (fluid attenuation inversion recovery), T1c (T1-contrasted), T1n (spin-lattice relaxation), T2w (spin-spin relaxation), and more [30]. Distinct combinations of modalities are employed in diverse clinical diagnoses, for example, myocardial infarction diagnosis involves LGE, cine, and T2 modalities, while brain tumor assessment utilizes T1c, T1n, T2w, and T2f modalities [16] [22] [3]. Different diseases require different information aggregation patterns from different modality combinations. Therefore, effective information aggregation from multi-modalities is of great clinical significance to quantify morphological and pathological changes, facilitating treatment planning and patient management.

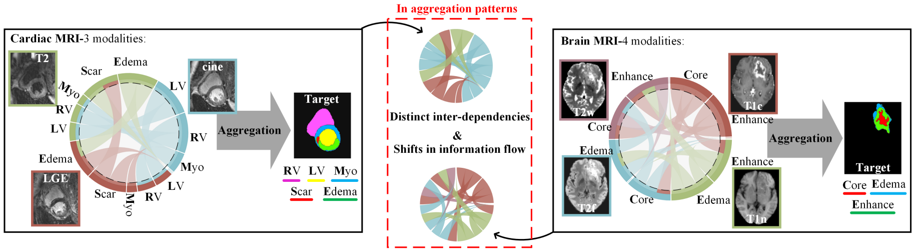
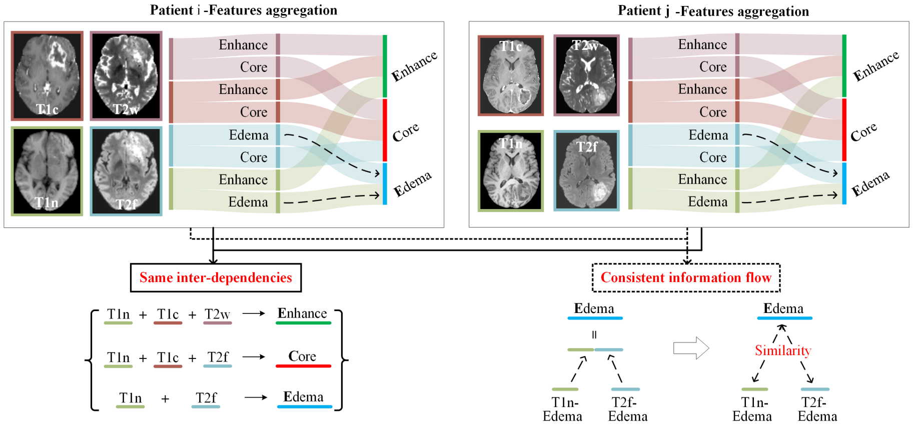


Fig. 1: Determining a unified multi-modality MRI aggregation pattern is still a challenge. Different diseases rely on distinct inter-dependencies and shifts in information flow among modalities.

However, determining a unified aggregation pattern for multi-modality MRI remains challenging. The information aggregation of multi-modality MRI follows a pipeline where information from different modalities flows in accordance with their inter-dependencies. A specific disease requires a specific aggregation pattern of various MRI modalities (Fig. 1). It can be found that the aggregations of different diseases rely on distinct inter-dependencies and shifts in information flow among modalities. Due to the different inter-dependencies and information flow in various aggregations, a unified aggregation pattern is difficult to define for the diseases in various scenarios.

Inter-dependencies and information flow are extensively investigated in the existing aggregation methods. Existing methods can be categorized into two groups based on whether the aggregation can be dynamically optimized. (1) Fixed. In these methods [33] [38] [11], the relevance of different modalities is defined as pre-operation according to the clinical experience. Hence, the aggregation performance relies on the quality of pre-defined relevance. The fixed aggregation requires these methods to redesign their frameworks when confronted with new scenarios. (2) Dynamic. In these methods [20] [18] [15] [23] [39] [30] [43], features from different modalities are concatenated into a whole, and rely on task labels to drive the network to automatically aggregate the features. Due to the lack of specific optimization targets for the inter-dependencies and information flow, the aggregation is implicit and unstable. As a result, a single framework operates differently when applied to various diseases with different modalities.



1. In the case of the same disease, multi-modality MRI exhibits the same inter-dependencies across different patients.

2. In the information flow, the aggregated information retains the consistent information from each modality.

Fig. 2: The properties of multi-modality MRI information aggregation: 1. In the case of the same disease, multi-modality MRI exhibits the same inter-dependencies across different patients. 2. In the information flow, the aggregated information retains the inherent details from each modality.

Motivated by the energy model, the inter-dependencies and information flow can be explicitly quantified to overcome the above challenges. Energy models capture dependencies among variables in various applications, including tasks like out-of-distribution detection [25], alignment in incremental learning [12], and structured prediction [1]. To quantify interdependence and information flow, we define energy based on the fundamental properties in both aspects(Fig. 2):

- **1.** In the case of the same disease, multi-modality MRI exhibits the same inter-dependencies across different patients' features. Consequently, the inter-dependencies of multi-modality MRI remain constant for the same scenario. In other words, within the same scenario, diverse patients' multi-modality features are aggregated in the same way.
- **2.** In the information flow, the aggregated feature preserves the consistent information derived from each modality. Those features for each disease are highly similar. Hence, there is consistency in the flow of information from multi-modality to aggregation.

According to the above properties, we propose a novel aggregation pattern, Energy-induced Explicit Propagation and Alignment (E^{2}PA), to enable the explicit quantification of multi-modality MRI fusion in different scenarios. In E^{2}PA:

(1) An energy-guided hierarchical fusion (EHF) uncovers the explicit quantification and optimization of inter-dependencies propagation by an attention-based enhancement and a hierarchical same energy among patients. (2) An energy-regularized space alignment (ESA) aligns the representations of different modalities by QR decomposition [10] and measures the consistency of information flow by energy minimization. The contributions are as described below.

- A novel aggregation prototype, Energy-induced Explicit Propagation and Alignment (E^{2}PA), explicitly quantifies aggregation from the inter-dependencies and information flow by the energy model. This fundamental property-based multi-modality aggregation pattern is adaptive to different medical scenarios directly and greatly boosts the performance of downstream diagnostic tasks.
- An energy-guided hierarchical fusion (EHF) enforces the optimization of inter-dependencies propagation of multi-modality MRI from attention-based representation and explicit quantification of the same energy among patients in hierarchical propagation. It establishes an equivalence between the measurement of inter-dependencies and the optimization of the energy to the aggregation.
- An energy-regularized space alignment (ESA) ensures the consistency of information flow in multi-modality aggregation by minimizing the energy on the factorized and aligned representation. It identifies a provable theory that guarantees the consistency of information in multi-modality aggregation.

2 Related works

Multi-modality MRI fusion To utilize information from multi-modality MRI assisting clinical diagnosis, extensive works focus on multi-modality MRI aggregation/fusion [2] [40] [32] [31]. The aggregation methods could be divided into:

(1) Task-driven (fixed). The fusion is constructed following the tasks related to different modalities [29]. [42] employs the different segmentation results to guide. In [33], the relevance of different pathology detection on multi-modality MRI is studied to select task-related modality MRI to aggregation. Following the relation among cardiac myocardium, scar, and edema. [37] set a coarse-to-fine way to fuse multi-modality CMR. Similarly, [21] sets a stack structure according to the order of various cardiac tissues segmentation for the information aggregation of multi-modality CMR. For different diseases, it makes the task-driven aggregation method requires retraining/redesign. Hence, task-driven aggregation has limited adaptability.

(2) Data-driven (dynamic). These methods weight multi-modality MRI according to the interdependence/similarity automatically. The weights of different modalities are calculated following the similarity by networks [28] [38]. To achieve adaptive weights for each modality, [30] designs an auto-weighted supervision mechanism to track the importance of the scar and edema segmentation. To search the inter-dependencies among multi-modalities, [13] and [43] conduct united adversarial learning to mine the correlation in liver tumor segmentation. In these methods, the aggregation results are evaluated by the downstream

segmentation/quantification tasks. The inter-dependencies and information flow lack specific optimization targets. Hence, a single framework operates differently when applied to various diseases with different modalities.

Energy-based models Energy-based model [14] assigns low energies to observed data-label pairs and high energies otherwise [8] [17] to maximize likelihood estimation. It has been applied to various aspects, including out-of-distribution sample detection [19] [25], structured prediction [1] [27], improving adversarial robustness [8] [35] and alignment in incremental learning [12]. In [44], the energy-based model transports source style to target style by implicit learning not the combination of normalized codes. [34] evaluates the performance of the energy-based model on domain adaptation. The above works all rely on the ability of energy functions to capture dependencies among variables. Hence, considering the same inter-dependencies and consistent information flow in multi-modality fusion, we propose to minimize the energy to explicitly quantify them.

3 Method

In our E^{2}PA (Fig. 3), the representations of multi-modalities are extracted by encoders respectively. EHF performs hierarchical attention-based fusion and explicitly quantifies the inter-dependencies among patients by defined energy. ESA measures the consistency of information flow on the aligned modality representation and aggregation results.

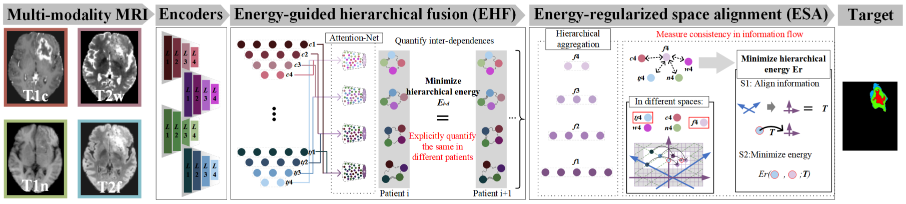


Fig. 3: E^{2}PA consists of: a) EHF uncovers the quantification and optimization of inter-dependencies among multi-modalities by hierarchical same energy among patients. b) ESA measures the consistency of information flow in multi-modality aggregation by the alignment on space factorization and the energy minimization.

3.1 EHF quantifies the inter-dependencies propagation

EHF adopts the attention-based fusion network and hierarchical energy quantification to explicitly quantify the inter-dependencies propagation of multi-modalities. For multi-modality MRI aggregation, a set of scanned MRI (with N patients) $\{(M_1^i, \dots, M_j^i, lab^i)\}_{i=1}^N$ denotes the j modalities MRI and the label of diagnosis target (lab). The hierarchical representations of each modality (Brain

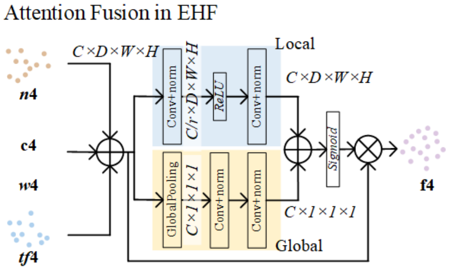


Fig. 4: The details of attention-based fusion network in EHF.

MRI for example $\{c1-c4\}, \{w1-w4\}, \{n1-n4\}, \{tf1-tf4\}$) are extracted by the encoders ($\{L1, L2, L3, L4\}$).

Attention-based fusion network captures the inter-dependencies by multi-scales attention. Inspired by the attentional feature fusion [4], attention-based fusion is established on multi-modality MRI. The multi-scale attention overcomes the difference on semantics and scales for the representation of inter-dependencies. On hierarchical representations, aggregation results $(f1, f2, f3, f4)$ are constructed by the attention fusion following the inter-dependencies among modalities (Fig.4). In hierarchical aggregation, the same attention network is utilized to aggregate the multi-modality MRI. Here, the fusion on $tf4$, $c4$, $n4$, and $w4$ is taken for example. Through the attention-based fusion, the representations of different modalities ($tf4$, $c4$, $n4$ and $w4$) are fused into $f4$. Firstly, the multi-modality representations with the size of $C \times D \times W \times H$ are added as a whole (A_f). Second, the fused representation is fed into local network to extract the local attention in the channel dimension. The A_f is extracted by the convolution and normalization layers with the size of $C/r \times D \times W \times H$. Then the ReLU, convolution, and normalization layers are applied to achieve the local channel-wise attention $(C \times D \times W \times H)$. Thirdly, the global attention is realized by a global-pooling and two convolution & normalization layers with the size of $C \times 1 \times 1 \times 1$. Fourth, the global and local attention is added and converted into sigmoid to achieve the attention in multi-modality MRI. Finally, the attention is applied to A_f for the attention-corrected fusion $(f4)$. The above fusion process is applied to hierarchical multi-modality MRI representations.

Energy function of EHF: To explicitly quantify the inter-dependencies propagation, EHF defines the relation between energy score and inter-dependencies. In the energy-based formulation, the target is to find an energy function i.e., $E_{i-d}(x, y)$ that gives the lowest energy to correct results and higher energy to other results (x is the inter-dependencies, y is the label). The aggregation must produce the value y^* for the smallest: $y^* = argminE_{i-d}(x, y)$. The joint probability of input x and label y can be estimated through the Gibbs distribution: $p(x, y) = exp(-E_{i-d}(x, y))/Z$, where $Z = \sum_x \sum_y exp(-E_{i-d}(x, y))$ is called the partition function that marginalizes over x and y. By marginalizing out y, the probability density for x can be achieved: $p(x) = \sum_y exp(-E_{i-d}(x, y))/Z$. Due to the difficulty of estimating Z, a free energy ($F(x)$) serves as the 'rationality' of the occurrence of the variable x:

$$F(x) = -log \sum exp(-E_{i-d}(x, y)) \tag{1}$$

To realize the quantification, we impose the proposition according to property (1):

- **Proposition 1.** *The same inter-dependencies among multi-modality MRI in different patients will make the energy $E_{i-d}(x, y)$ be equal for any instance.*

Precisely, the label y, which cannot be quantified, is converted to the inter-dependencies of another patient (x'). x and x' are generated by encoders and attention fusions ($(L + at)$ and $(L + at)'$). This situation follows the *latent variable* [14]. Due to the different inputs of the encoder, the inherent information in the aggregation is the inter-dependencies. Hence, the energy function is updated to $E_{i-d}(x, x'; (L+at), (L + at)')$. To avoid the collapse in energy optimization on two variables [14], EHF takes two ways: i) Fixing one variable in optimization. The $(L + at)'$ is fixed to generate fixed x' as inter-dependencies. With iterative optimization, the measurement of inter-dependencies will be improved and achieve a stable value finally. ii) Enlarging energy to incorrect answers. The negative likelihood from probabilistic modeling is utilized to regularize. Hence, the energy of inter-dependencies can be formulated as:

$$\begin{aligned}&\mathcal{L}_{i-d}(x, x'; (L + at), (L + at)') = E_{i-d}(x, x'; (L + at),\\ &(L + at)') + log \sum exp(-E_{i-d}(x, \tilde{x}'; (L + at), \widetilde{(L + at)}'))\end{aligned} \tag{2}$$

,where $\tilde{x}'$ is the incorrect answer generated by $\widetilde{(L + at)}'$. It can be simplified:

$$\mathcal{L}_{i-d}(x, x'; (L + at), (L + at)') = \sum_{L1-L4} E_{i-d}(x, x'; (L + at), (L + at)') - F(x) \tag{3}$$

By this energy calculation on hierarchical aggregation, the inter-dependencies in multi-modalities can be explicitly quantified through iterative optimizations.

3.2 ESA ensures the consistency in information flow

ESA ensures the consistency of information flow in multi-modality aggregation by two steps: align information and minimize energy (Fig. 5). EHF focuses on the same inter-dependencies will motivate the encoders and attention fusion ($(L + at)$) to prefer the same information in different patients and ignore the consistency in information flow.

Information Alignment. Since the information of different modalities is in different spaces, the measurement of consistency can not be directly on the information flow (proof in Supplementary). Hence, information alignment is necessary before measurement. It can be formulated to:

$$\mathcal{F}_{align}(f, c, l, t) = \sum_{i=1}^{4} ci, li, ti \overset{T}{\rightarrow} fi \tag{4}$$

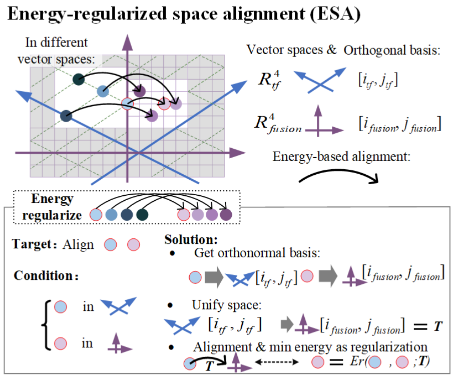

Energy-regularized space alignment (ESA)
In different vector spaces:
Vector spaces & Orthogonal basis:
R^4_{tf}
$[i_{tf}, j_{tf}]$
R^4_{fusion}
$[i_{fusion}, j_{fusion}]$
Energy-based alignment:
Energy regularize
Target: Align
Condition:
in
in
Solution:
Get orthonormal basis:
$[i_{tf}, j_{tf}]$
$[i_{fusion}, j_{fusion}]$
Unify space:
$[i_{tf}, j_{tf}]$
$[i_{fusion}, j_{fusion}] = T$
Alignment & min energy as regularization
T
= Er(, ;T)

Fig. 5: The details of the alignment process in ESA. The information of multi-modalities in different vector spaces is aligned into the same space for effective fusion.

where T is the transformation function of spaces. The information of different modalities and aggregation is represented as a $m \times n$ matrix. The matrix can be factorized to one orthogonal matrix Q and one upper triangular matrix R (QR decomposition [10]):

$$f4 = Q^4_{fusion} @ R^4_{fusion}, \quad tf4 = Q^4_{tf} @ R^4_{tf} \tag{5}$$

where @ is the matrix multiplication operation, the dimension of Q is a $m \times n$ and R is a $n \times n$ matrix. The Q, is the space of R, which can be viewed as the orthogonal basis of each modality representation. Hence, the information alignment turns into the problem solving T on the Q of fusion results ($f4$) and multi-modality MRI ($tf4, c4, n4, w4$). To simplify the solution process, $f4$ and $tf4$ are taken example: $Q^4_{tf} \xrightarrow{T} Q^4_{fusion}$. To find the value of T, the process is converted to an equation (viewing T as a matrix):

$$Q^4_{tf} @ T = Q^4_{fusion} \quad Q^4_{tf} @ {Q^4_{tf}}^{-1} @ Q^4_{fusion} = Q^4_{fusion}. \tag{6}$$

Hence, T is equal to ${Q^4_{cine}}^{-1} @ Q^4_{fusion}$. Since the Q^4_{fusion} is known, the target is to find ${Q^4_{tf}}^{-1}$. Since the property of orthogonal matrix on Q:

$$Q^4_{tf} @ {Q^4_{tf}}^{-1} = I, Q^4_{tf} @ {Q^4_{tf}}^{T} = I, \tag{7}$$

it can easily find ${Q^4_{tf}}^{T} = {Q^4_{tf}}^{-1}$. Hence, the value of T can be obtained: $T = {Q^4_{tf}}^{T} @ Q^4_{fusion}$. Since the upper triangular matrix R^4_{tf} is corresponding to Q^4_{tf}, the alignment requires to be performed on it ($R^4_{tf} @ T = {R^4_{tf}}'$). Finally, the aligned information is:

$$tf4_a = Q^4_{fusion} @ {R^4_{tf}}'. \tag{8}$$

Since the $f4$ and $tf4$ are aligned into the same space, the measurement of consistency can be conducted directly on $f4$ and $tf4_a$. The same operation is performed on hierarchical multi-modality information.

Energy function of ESA: With the aligned information, ESA introduces an energy function to measure the information consistency between the aggregation and multi-modality MRI. It is based on the property (2):

- **Proposition 2.** *The information in multi-modality MRI is maintained in the aggregated results.* The aggregation needs to have low energy $E_r(x, y)$ with multi-modality representations.

The above situation follows the implicit regression [14] in energy-based model: In the aggregation, the multi-modality MRI representations (multiple answers) are equally good. Simplify, the dependency between the aggregation result $f4$ and multi-modality representations $tf4_a, c4_a, n4_a, w4_a$ cannot be formulated as a mapping from $f4$ to $tf4_a$, $c4_a$, $n4_a$, or$w4_a$ directly. Hence, ESA models the consistency and according it to design energy function:

$$E_r(f4, tf4_a, c4_a, n4_a, w4_a) = \sum_{i}^{tf,c,n,w} \left\| G_{at}(f4) - G_L i4_a \right\|^2 . \tag{9}$$

where G_{at} and G_L are the model functions to achieve $f4$ and $tf4_a$. Through minimization of $E_r()$, the information from different modalities will be maintained under this constraint. The regularization can be formulated as:

$$\mathcal{L}_r = \sum_{i}^{1,2,3,4} E_r(fi, \mathcal{F}_{align}(fi, ci), \mathcal{F}_{align}(fi, li), \mathcal{F}_{align}(fi, ti)). \tag{10}$$

Compared with $\mathcal{L}_{i-d}$, which guides the same inter-dependencies from different patients, $\mathcal{L}_r$ constrains the consistency in information flow in the same patient. Through the minimization of two energy functions on hierarchical aggregation, $\mathcal{L} = \mathcal{L}_{i-d} + \mathcal{L}_r$, the inter-dependencies and consistency are explicitly quantified. With different downstream tasks, $\mathcal{L}$ can be combined with the optimization of the target directly.

Table 1: The description of three public datasets (seg-segmentation, class-classification). Here MRNet is a public dataset.

Name	Modality	Target	Train	Test	Total
MyoPS	cine,LGE,T2	Seg	15	10	25
BraTS	T1c, T1n, T2f, T2w	Seg	834	417	1251
MRNet	Sagittal T2,Axial PD, Coronal T1	Class	753	377	1130

4 Experiments

We evaluate E^{2}PA on various multi-modality MRI scenarios, including: cardiac pathology segmentation (MyoPS dataset [16]), brain tumor segmentation (BraTS2021 dataset [22]), and detection of anterior cruciate ligament tears (MRNet dataset [3]). The division of training and testing data is shown in Tab. 1. Our E^{2}PA is based on Pytorch with the Adam optimization. The encoders are same as U-net [24]. All experiments are performed on a single NVIDIA TITAN RTX GPU. The metrics of Dice [6] and Area Under Curve (AUC) are utilized for segmentation and classification performance evaluation, respectively. More details of experiments are presented in Supplementary.

4.1 Main results

For different downstream tasks, various state-of-the-art multi-modality MRI methods are adopted for comparison. **For segmentation task (MyoPS and BraTS)**: AWSNet [30], MyoPS-Net [23], NestedFormer [36], HyperDense-Net [7], MAML [41], and MMSNet [45]. **For classification task (MRNet)**: TransMed [5], MRNet [3], ELNet [26], and MRPyrNet [9].

Table 2: The quantitative results on BraTS and MyoPS reveal the superior ability of our E^{2}PA on multi-modality MRI segmentation.

Methods	BraTS%				MyoPS%					
	TC	ET	WT	AVG	scar	edema	LV	RV	MYO	AVG
HyperDense-Net	84.1	80.3	87.1	83.8	57.3	68.6	93.1	87.8	77.9	76.9
MAML	86.9	85.4	89.9	87.4	61.1	70.9	93.3	89.9	79.3	78.9
MMSNet	83.3	82.1	88.4	84.6	60.5	73.6	94.1	89.3	83.8	80.3
AWSNet	87.0	86.6	92.8	88.8	61.1	72.3	92.8	88.2	81.4	79.2
NestedFormer	88.4	85.1	91.3	88.3	62.0	73.1	93.9	89.9	84.7	80.7
MyoPS-Net	90.1	81.1	89.3	86.8	63.4	**74.0**	94.0	**92.0**	86.1	81.9
Ours	**91.0**	**87.3**	**93.5**	**90.6**	**64.7**	73.9	**94.4**	91.1	**87.0**	**82.2**

BraTS: The experimental results for brain tumor segmentation from four modalities MRI (T1c, T1n, T2f, T2w) evaluate the superiority of aggregation. The targets contain tumor core (TC), enhancing tumor (ET) and whole tumor (WT). **Quantitative:** It can be observed from Tab. 2 that HyperDense-Net achieves the lowest performance on each tissue (84.1%, 80.3%, and 87.1%). This indicates that the specifically designed fusion modules in the network will be more effective than the direct feature concatenating (HyperDense-Net). Our E^{2}PA achieves the highest Dice score 90.6%, and this indicates that the proposed multi-modality MRI aggregation is superior to others. The lower Dice score on ET than on other targets (WT, TC) indicates the ET is difficult to segment. The highest Dice score of our E^{2}PA on ET (87.3%) proves the ability for multi-modality aggregation. **Qualitative:** The qualitative analysis on brain MRI evaluates the significance on various regions. From Fig. 6, it can be found that: The core of tumor (red) can be well detected in all methods. This indicates that these multi-modality methods could recognize the significant tumor feature. For the enhance (green) and edema (blue) regions of tumor, our E^{2}PA achieves the best performance. This indicates that the tumor information is well aggregated. The balance performance on different tumor regions also proves that E^{2}PA can obtain the representations from multi-modality MRI better than others.

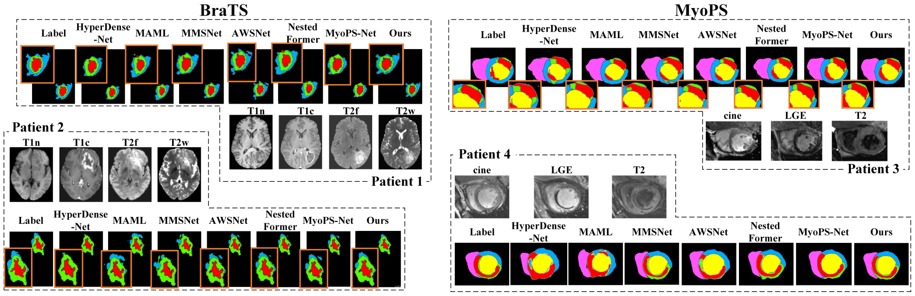


Fig. 6: The framework of FTP: (a)Task-aware memory network stores the mapping (task identity to model weights) by the hypernetwork to avoid catastrophic forgetting. (b) Adaptive prototype matching aggregates the prototypes for continual optimization on heterogeneous streams.

MyoPS: The experimental results The targets contain scar, edema, left ventricle (LV), right ventricle (RV) and myocardium (MYO). **Quantitative:** The quantitative results on the segmentation from multi-modality MRI (LGE, T2, cine) evaluate the superiority of E^{2}PA. For the regions of scar, LV, MYO, and average, our E^{2}PA achieves the best Dice score (64.7%, 94.4%, 87% and 82.2%). This indicates the aggregation of E^{2}PA is effective for cardiac MRI. In the regions of edema and RV, our E^{2}PA achieves similar performance with MyoPS-Net. MyoPS-Net relies on various combinations in representation and aggregation for each target region segmentation. The similar results also prove that E^{2}PA is effective for multi-modality MRIaggregation. **Qualitative:** The qualitative analysis on cardiac MRI evaluates the significance on different regions. In Fig. 6, the edema (green) and scar (red) regions are more difficult to segment than LV (yellow) & RV (pink) & MYO (blue). E^{2}PA achieves fine-grained segmentation on these regions. This indicates its superiority on achieving target-related information from multi-modality. Since the scar and edema are on the MYO, the best segmentation of E^{2}PA also proves that the significance of different regions in different modalities is well aggregated.

MRNet dataset: Our E^{2}PA achieves superior classification performance on multi-modality knee MRI. The state-of-the-art classification methods for the classification of anterior cruciate ligament tears on MRNet dataset are compared, including TransMed [5], MRNet [3], ELNet [26], and MRPyrNet [9]. Our E^{2}PA

Table 3: The superior performance of E^{2}PA on classification from multi-modality knee MRI. MRPyrNet is embedded into MRNet (method) and ELNet respectively (Abn-Abnormal, Men-Meniscal).

AUC	MRNet	ELNet	MRPyrNet (MRNet)	MRPyrNet (ELNet)	TransMed	**Ours**
Abn	93.0%	93.7%	93.1%	94.0%	95.8%	**97.8%**
ACL	95.1%	94.9%	96.0%	95.7%	96.3%	**97.5%**
Men	83.3%	86.8%	89.3%	89.1%	92.3%	**94.4%**

achieves 2%, 1.2%, and 2.1% higher AUC in the classification of abnormal, ACL tear, and meniscal tear(Tab. 3). This indicates that the energy-induced aggregation of E^{2}PA provides a better representation of multi-modality MRI for the classification task than other multi-modality analysis methods. Compared with the ELNet and MRPyrNet, which utilize clinical knowledge to locate anomalies as priors, our E^{2}PA achieves better performance through the optimization of the energy functions and the classification label guided loss.

Table 4: The ablation study indicates the contributions of different modules in E^{2}PA.

EHF		ESA		Dice
Attention Fusion	$\mathcal{L}_{i-d}$	Alignment	$\mathcal{L}_r$	
✓	✓	✓	✓	90.6%
	✓	✓	✓	88.1%
✓	✓		✓	81.0%
✓		✓	✓	86.5%
✓	✓	✓		83.1%

4.2 Model analysis

To further analyze the proposed E^{2}PA, we design ablation study, inter-dependencies analysis, and consistency of information flow analysis on BraTS and MyoPS datasets.

Ablation study. To evaluate the contribution of each module in E^{2}PA, different ablation strategies are designed for comparison on BraTS dataset. E^{2}PA contains EHF (attention fusion and $\mathcal{L}_{i-d}$) and ESA (Alignment and $\mathcal{L}_r$). It can be found that each module of E^{2}PA is effective for multi-modality MRI segmentation (Tab. 4). The attention fusion module and $\mathcal{L}_{i-d}$ bring 2.5% and 4.1% improvements on Dice score. This indicates that the energy-guided inter-dependencies quantification can further improve the aggregation. The alignment and $\mathcal{L}_r$ brings 9.6% and 7.5% improvements. The QR decomposition-based alignment and energy-regularized inherent information can further improve the consistency of information flow.

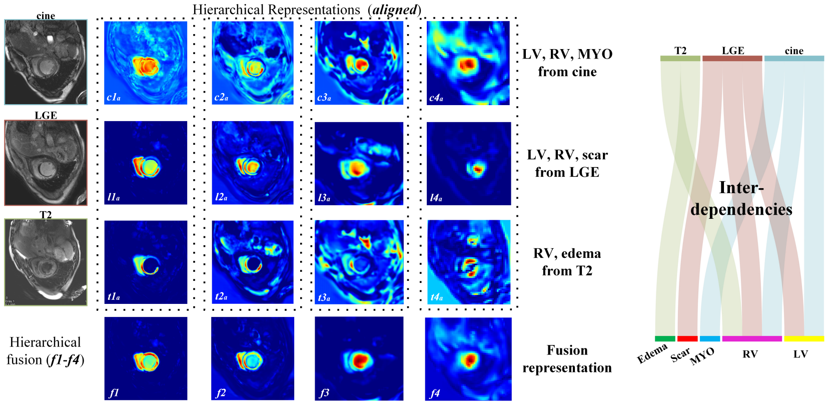


Fig. 7: The evaluation of the quantification of inter-dependencies. The visual representations reveal the inter-dependencies in aggregation. The hierarchical representations are aligned.

Inter-dependencies. To analyze the quantification of inter-dependencies, the representations of different modalities and aggregation are visualized. From Fig. 7, different regions of interest are shown in various modalities of cardiac MRI. In cine, the information of LV, RV, and MYO are obvious in hierarchical representations. In LGE, RV & LV & scar regions can be observed. In T2, RV and edema regions are obvious. The hierarchical aggregations of our E^{2}PA contain all the target-related regions from multi-modality MRI. Hence, the inter-dependencies among the three modalities can be quantified that: i) LV segmentation is related to LGE & cine; ii) RV segmentation is related to LGE & cine & T2; iii) MYO segmentation is related to cine only; iv) scar information is from LGE; v) edema information is from T2. This also reveals that E^{2}PA has the ability to quantify the inter-dependencies among multi-modality MRI. More inter-dependencies are shown in Supplementary.

Consistency of information flow. To analyze the consistency of information flow in the aggregation, the residuals between aggregation representations ($f1 - f4$) and multi-modality representations ($c1 - c4$, $l1 - l4$, $t1 - t4$) & aligned representations ($c1_a - C4_a$, $l1_a - l4_a$, $t1_a - t4_a$). From Fig. 8, the residual representation between aggregation and LGE is that: i) The less LV & RV & scar representations in the residual reveals that this information in aggregation is from LGE. The consistent information is maintained by E^{2}PA. ii) The aligned representation $l1_a$ has less residual with aggregation $f1$ than $l1$. This reveals that the alignment in ESA promotes the measurement of consistency in information flow in aggregation.

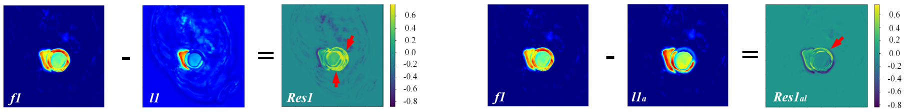


Fig. 8: The residual reveals the consistency of information flow in aggregation, and the essential alignment in E^{2}PA.

5 Conclusion

In this paper, we propose an energy-induced Propagation and Alignment (E^{2}PA) to explicitly quantify the inter-dependencies by hierarchical same energy among patients (EHF) and measure the consistency of information flow (ESA). The novel aggregation prototype optimizes the properties of multi-modality MRI aggregation to adapt to different scenarios. Through the extensive experiments on 3 public multi-modality MRI datasets (with different modality combinations and tasks), the superiority of E^{2}PA can be demonstrated from the comparison with state-of-the-art methods. It will greatly boost the performance of downstream clinical diagnostic tasks.

Acknowledgments: This work was supported by Ministry of Education Tier 1 Start up grant, NUS, Singapore (A-8001267-01-00). We thank the Big Data Computing Center of Southeast University for providing the facility support.

References

1. Bakhtin, A., Deng, Y., Gross, S., Ott, M., Ranzato, M., Szlam, A.: Residual energy-based models for text. The Journal of Machine Learning Research **22**(1), 1840–1880 (2021) 3, 5
2. Baltrušaitis, T., Ahuja, C., Morency, L.P.: Multimodal machine learning: A survey and taxonomy. IEEE transactions on pattern analysis and machine intelligence **41**(2), 423–443 (2018) 4
3. Bien, N., Rajpurkar, P., Ball, R.L., Irvin, J., Park, A., Jones, E., Bereket, M., Patel, B.N., Yeom, K.W., Shpanskaya, K., et al.: Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of mrnet. PLoS medicine **15**(11), e1002699 (2018) 2, 10, 11
4. Dai, Y., Gieseke, F., Oehmcke, S., Wu, Y., Barnard, K.: Attentional feature fusion. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 3560–3569 (2021) 6
5. Dai, Y., Gao, Y., Liu, F.: Transmed: Transformers advance multi-modal medical image classification. Diagnostics **11**(8), 1384 (2021) 10, 11
6. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26**(3), 297–302 (1945) 10
7. Dolz, J., Gopinath, K., Yuan, J., Lombaert, H., Desrosiers, C., Ayed, I.B.: Hyperdense-net: a hyper-densely connected cnn for multi-modal image segmentation. IEEE transactions on medical imaging **38**(5), 1116–1126 (2018) 10
8. Du, Y., Mordatch, I.: Implicit generation and modeling with energy based models. Advances in Neural Information Processing Systems **32** (2019) 5

9. Dunnhofer, M., Martinel, N., Micheloni, C.: Improving mri-based knee disorder diagnosis with pyramidal feature details. In: Medical Imaging with Deep Learning. pp. 131–147. PMLR (2021) 10, 11
10. Gander, W.: Algorithms for the qr decomposition. Res. Rep **80**(02), 1251–1268 (1980) 4, 8
11. Havaei, M., Guizard, N., Chapados, N., Bengio, Y.: Hemis: Hetero-modal image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19. pp. 469–477. Springer (2016) 2
12. Joseph, K., Khan, S., Khan, F.S., Anwer, R.M., Balasubramanian, V.N.: Energy-based latent aligner for incremental learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7452–7461 (2022) 3, 5
13. Kamnitsas, K., Baumgartner, C., Ledig, C., Newcombe, V., Simpson, J., Kane, A., Menon, D., Nori, A., Criminisi, A., Rueckert, D., et al.: Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In: Information Processing in Medical Imaging: 25th International Conference, IPMI 2017, Boone, NC, USA, June 25-30, 2017, Proceedings 25. pp. 597–609. Springer (2017) 4
14. LeCun, Y., Chopra, S., Hadsell, R., Ranzato, M., Huang, F.: A tutorial on energy-based learning. Predicting structured data **1**(0) (2006) 5, 7, 9
15. Li, F., Li, W.: Dual-path feature aggregation network combined multi-layer fusion for myocardial pathology segmentation with multi-sequence cardiac mr. In: Myocardial Pathology Segmentation Combining Multi-Sequence Cardiac Magnetic Resonance Images: First Challenge, MyoPS 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings 1. pp. 146–158. Springer (2020) 2
16. Li, L., Wu, F., Wang, S., Luo, X., Martín-Isla, C., Zhai, S., Zhang, J., Liu, Y., Zhang, Z., Ankenbrand, M.J., et al.: Myops: A benchmark of myocardial pathology segmentation combining three-sequence cardiac magnetic resonance images. Medical Image Analysis **87**, 102808 (2023) 2, 10
17. Li, S., Du, Y., van de Ven, G., Mordatch, I.: Energy-based models for continual learning. In: Conference on Lifelong Learning Agents. pp. 1–22. PMLR (2022) 5
18. Li, W., Wang, L., Qin, S.: Cms-unet: Cardiac multi-task segmentation in mri with a u-shaped network. In: Myocardial Pathology Segmentation Combining Multi-Sequence Cardiac Magnetic Resonance Images: First Challenge, MyoPS 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings 1. pp. 92–101. Springer (2020) 2
19. Liu, W., Wang, X., Owens, J., Li, Y.: Energy-based out-of-distribution detection. Advances in neural information processing systems **33**, 21464–21475 (2020) 5
20. Liu, Y., Zhang, M., Zhan, Q., Gu, D., Liu, G.: Two-stage method for segmentation of the myocardial scars and edema on multi-sequence cardiac magnetic resonance. In: Myocardial Pathology Segmentation Combining Multi-Sequence Cardiac Magnetic Resonance Images: First Challenge, MyoPS 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings 1. pp. 26–36. Springer (2020) 2
21. Martín-Isla, C., Asadi-Aghbolaghi, M., Gkontra, P., Campello, V.M., Escalera, S., Lekadir, K.: Stacked bcdu-net with semantic cmr synthesis: Application to myocardial pathology segmentation challenge. In: Myocardial Pathology Segmentation Combining Multi-Sequence Cardiac Magnetic Resonance Images: First Challenge, MyoPS 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings 1. pp. 1–16. Springer (2020) 4

22. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014) 2, 10
23. Qiu, J., Li, L., Wang, S., Zhang, K., Chen, Y., Yang, S., Zhuang, X.: Myops-net: Myocardial pathology segmentation with flexible combination of multi-sequence cmr images. Medical Image Analysis **84**, 102694 (2023) 2, 10
24. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015) 10
25. Tonin, F., Pandey, A., Patrinos, P., Suykens, J.A.: Unsupervised energy-based out-of-distribution detection using stiefel-restricted kernel machine. In: 2021 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2021) 3, 5
26. Tsai, C.H., Kiryati, N., Konen, E., Eshed, I., Mayer, A.: Knee injury detection using mri with efficiently-layered network (elnet). In: Medical Imaging with Deep Learning. pp. 784–794. PMLR (2020) 10, 11
27. Tu, L., Gimpel, K.: Learning approximate inference networks for structured prediction. arXiv preprint arXiv:1803.03376 (2018) 5
28. Valverde, S., Cabezas, M., Roura, E., González-Villà, S., Pareto, D., Vilanova, J.C., Ramió-Torrentà, L., Rovira, À., Oliver, A., Lladó, X.: Improving automated multiple sclerosis lesion segmentation with a cascaded 3d convolutional neural network approach. NeuroImage **155**, 159–168 (2017) 4
29. Vielzeuf, V., Lechervy, A., Pateux, S., Jurie, F.: Centralnet: a multilayer approach for multimodal fusion. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops. pp. 0–0 (2018) 4
30. Wang, K.N., Yang, X., Miao, J., Li, L., Yao, J., Zhou, P., Xue, W., Zhou, G.Q., Zhuang, X., Ni, D.: Awsnet: An auto-weighted supervision attention network for myocardial scar and edema segmentation in multi-sequence cardiac magnetic resonance images. Medical Image Analysis **77**, 102362 (2022) 2, 4, 10
31. Wang, L., Shi, F., Gao, Y., Li, G., Gilmore, J.H., Lin, W., Shen, D.: Integration of sparse multi-modality representation and anatomical constraint for isointense infant brain mr image segmentation. NeuroImage **89**, 152–164 (2014) 4
32. Wang, L., Shi, F., Lin, W., Gilmore, J.H., Shen, D.: Automatic segmentation of neonatal images using convex optimization and coupled level sets. NeuroImage **58**(3), 805–817 (2011) 4
33. Xiao, X., Zhao, J., Li, S.: Task relevance driven adversarial learning for simultaneous detection, size grading, and quantification of hepatocellular carcinoma via integrating multi-modality mri. Medical Image Analysis **81**, 102554 (2022) 2, 4
34. Xie, B., Yuan, L., Li, S., Liu, C.H., Cheng, X., Wang, G.: Active learning for domain adaptation: An energy-based approach. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 8708–8716 (2022) 5
35. Xie, J., Lu, Y., Zhu, S.C., Wu, Y.: A theory of generative convnet. In: International Conference on Machine Learning. pp. 2635–2644. PMLR (2016) 5
36. Xing, Z., Yu, L., Wan, L., Han, T., Zhu, L.: Nestedformer: Nested modality-aware transformer for brain tumor segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 140–150. Springer (2022) 10

37. Zhai, S., Gu, R., Lei, W., Wang, G.: Myocardial edema and scar segmentation using a coarse-to-fine framework with weighted ensemble. In: Myocardial Pathology Segmentation Combining Multi-Sequence Cardiac Magnetic Resonance Images: First Challenge, MyoPS 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings 1. pp. 49–59. Springer (2020) 4
38. Zhang, D., Huang, G., Zhang, Q., Han, J., Han, J., Wang, Y., Yu, Y.: Exploring task structure for brain tumor segmentation from multi-modality mr images. IEEE Transactions on Image Processing **29**, 9032–9043 (2020) 2, 4
39. Zhang, J., Xie, Y., Liao, Z., Verjans, J., Xia, Y.: Efficientseg: A simple but efficient solution to myocardial pathology segmentation challenge. In: Myocardial Pathology Segmentation Combining Multi-Sequence Cardiac Magnetic Resonance Images: First Challenge, MyoPS 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings 1. pp. 17–25. Springer (2020) 2
40. Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., Shen, D.: Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. NeuroImage **108**, 214–224 (2015) 4
41. Zhang, Y., Yang, J., Tian, J., Shi, Z., Zhong, C., Zhang, Y., He, Z.: Modality-aware mutual learning for multi-modal medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. pp. 589–599. Springer (2021) 10
42. Zhang, Z., Liu, C., Ding, W., Wang, S., Pei, C., Yang, M., Huang, L.: Multi-modality pathology segmentation framework: application to cardiac magnetic resonance images. In: Myocardial Pathology Segmentation Combining Multi-Sequence Cardiac Magnetic Resonance Images: First Challenge, MyoPS 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings 1. pp. 37–48. Springer (2020) 4
43. Zhao, J., Li, D., Xiao, X., Accorsi, F., Marshall, H., Cossetto, T., Kim, D., McCarthy, D., Dawson, C., Knezevic, S., et al.: United adversarial learning for liver tumor segmentation and detection of multi-modality non-contrast mri. Medical image analysis **73**, 102154 (2021) 2, 4
44. Zhao, Y., Chen, C.: Unpaired image-to-image translation via latent energy transport. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 16418–16427 (2021) 5
45. Zhou, T., Canu, S., Vera, P., Ruan, S.: 3d medical multi-modal segmentation network guided by multi-source correlation constraint. In: 2020 25th International Conference on Pattern Recognition (ICPR). pp. 10243–10250. IEEE (2021) 10